NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

FORECAST ERROR METRICS FOR NAVY INVENTORY MANAGEMENT PERFORMANCE

by

Kenneth J. Jackson

March 2011

Thesis Advisor:	Ronald D. Fricker, Jr.
Second Reader:	Patrick A. R. Burson

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2011	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE
Forecast Error Metrics for Navy Inventory Management Performance

5. FUNDING NUMBERS

6. AUTHOR(S) Kenneth J. Jackson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Inventory Control Point, Mechanicsburg, PA

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol Number ________N/A

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

This research establishes metrics for determining overall Navy secondary inventory forecasting accuracy when compared to actual demands at the Naval Inventory Control Point (NAVICP). Specifically, two performance metrics are introduced: the average performance index (API) and the median absolute deviation performance index (MPI). API measures forecasting accuracy of secondary inventory when compared against demand or forecast performance over a four-quarter period. MPI measures the quarterly variability of forecast errors over the same period. The API and MPI metrics allow for the identification of poorly forecasted NAVICP secondary inventory items. The metrics can be applied to entire inventories or subsets of items based on type, demand, or cost. In addition, the API metric can be used to show overall inventory performance, providing NAVICP with a graphical means to assess forecasting performance improvements (or degradations) over time.

The new forecasting accuracy methods developed in this research will allow the Navy to continually gauge the overall health of their inventory management practices and provide a method for improving forecasting accuracy. Additionally, they will assist NAVICP in complying with DoD directives that require NAVICP to monitor and continually develop improvements to inventory management practices.

14. SUBJECT TERMS
Inventory, Demand, Forecast, Forecast Error, Inventory Management, MAD, CHURN

15. NUMBER OF PAGES 65

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

FORECAST ERROR METRICS FOR NAVY INVENTORY MANAGEMENT PERFORMANCE

Kenneth J. Jackson
Lieutenant Commander, United States Navy
B.B.A., University of Cincinnati, 1998

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 2011

Author: Kenneth J. Jackson

Approved by: Ronald D. Fricker, Jr.
Thesis Advisor

Patrick A. R. Burson
Second Reader

Robert F. Dell
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This research establishes metrics for determining overall Navy secondary inventory forecasting accuracy when compared to actual demands at the Naval Inventory Control Point (NAVICP). Specifically, two performance metrics are introduced: the average performance index (API) and the median absolute deviation performance index (MPI). API measures forecasting accuracy of secondary inventory when compared against demand or forecast performance over a four quarter period. MPI measures the quarterly variability of forecast errors over the same period.

The API and MPI metrics allow for the identification of poorly forecasted NAVICP secondary inventory items. The metrics can be applied to entire inventories or subsets of items based on type, demand, or cost. In addition, the API metric can be used to show overall inventory performance, providing NAVICP with a graphical means to assess forecasting performance improvements (or degradations) over time.

The new forecasting accuracy methods developed in this research will allow the Navy to continually gauge the overall health of their inventory management practices and provide a method for improving forecasting accuracy. Additionally, they will assist NAVICP in complying with DoD directives that require NAVICP to monitor and continually develop improvements to inventory management practices.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

API	Annual Performance Metric
COG	Cognizance Code
DoD	Department of Defense
GAO	Government Accountability Office
LASE	Lead-time Adjusted Symmetric Error
MAD	Median Absolute Deviation
MAPE	Mean Absolute Percentage Error
MPI	Median Absolute Deviation Performance Index
MSE	Mean Square Error
NAVICP-M	Naval Inventory Control Point, Mechanicsburg
NAVICP-P	Naval Inventory Control Point, Philadelphia
Navy ERP	Navy Enterprise Resource Planning system
NIIN	National Item Identification Number
RTAT	Repair Turn-Around Time
PI	Performance Indicator
PLT	Procurement Lead-Time
PRLT	Production Lead-Time
PRTAT	Process Repair Turn-Around Time
SD	Standard Deviation
UICP	Uniform Inventory Control Program
USD	U.S. Dollars

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

This research establishes new metrics for quantifying Navy secondary inventory forecasting accuracy when compared to actual demands. Specifically, two performance metrics are introduced: the average performance index (API) and the median absolute deviation (MAD) performance index (MPI). API measures forecasting accuracy of secondary inventory when compared against demand performance over a one-year period. MPI is a measure of the variability of forecast errors over the same period.

Once calculated, the API and MPI metrics are utilized in the development of an overall forecasting accuracy method. Together, these methods allow for the identification of National Item Identification Numbers (NIINs) with poorly performing forecasts as well as the assessment of the overall performance of all NAVICP secondary inventory items. Additionally this method provides NAVICP with a graphical means of illustrating overall forecasting performance from quarter to quarter that shows trends in forecasting performance over time.

The new forecasting accuracy methods developed in this research will allow the Navy to continually gauge the overall health of inventory management practices and provide a method for improving forecasting accuracy. Additionally, the methods will assist NAVICP in complying with DoD directives which require it to monitor and continually develop improvements to inventory management practices (DoD, 2004). Finally, improvements in forecasting accuracy should reduce excess inventory and thus save taxpayer dollars.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

“If I have seen further it is only by standing on the shoulders of Giants.” –Sir Isaac Newton

I would like to express my heartfelt gratitude to my thesis advisor, Professor Ron Fricker. Thank you for your outstanding guidance, tireless dedication, and patience throughout my thesis studies. I would also like to thank both CDRs Patrick A. R. Burson and Walt DeGrange, whose professional knowledge and mentorship have always been valued and appreciated.

Lastly, I want to thank my family and friends for their unwavering support throughout my studies. Your constant words of encouragement and belief in me made this all possible. I am truly blessed to have such amazing people in my life and could not have done it without you.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

The U.S. Navy performs inventory management on $18.7 billion worth of secondary inventory. Secondary inventory consists of reparable components, consumable repair parts, bulk items and materiel, subsistence, and expendable end items, including clothing and other personal gear (DoD, 2003). In an attempt to balance the needs for meeting warfighter readiness, coupled with competing resources and good stewardship, the Navy must actively manage these requirements on secondary inventory to effectively achieve inventory efficiency. The Department of Defense (DoD) Supply Chain Management Regulation requires all services to actively pursue efficient and effective materiel support at least cost (DoD, 2004).

Currently, there is no mechanism in place to determine overall inventory efficiency for the U.S. Navy. Specifically, the Navy has no established metrics to measure the accuracy or performance of its inventory management system. The Navy does not have a formal method to determine how well forecasted demand matches actual demand in the Navy’s secondary inventory. Because no methods or metrics have been established or utilized, the Navy has no way to determine overall inventory management effectiveness. Therefore, the Navy is unable to track if improvements are being made to forecast accuracy or if a deficiency persists.

B. GOVERNMENT ACCOUNTABILITY OFFICE (GAO)

In 2007, the United States Government Accountability Office (GAO) conducted an audit of the Navy inventory system to determine its inventory management effectiveness. Among its findings, the GAO determined that the Navy held excess levels of inventory. The GAO identified $7.5 billon exceeded current requirements of the $18.7 billion in Navy secondary inventory for the years 2004 to 2007. About half of the $7.5 billion was identified as potential excess. The GAO also reported the Navy had not established the cost efficiency of its inventory management system over the 4-year period, that the Navy's demand forecasting effectiveness is limited as requirements for items may change frequently after purchase decisions are made, and the Navy failed to adjust management practices in response to addressing unpredictability in demand (GAO, 2008).

Although the Navy established metrics to measure warfighter support, it lacked metrics and goals for tracking and measuring the cost efficiency of its inventory management. Specifically, the Navy has not analytically evaluated the effectiveness of its demand forecasting. Problems with demand forecasting include inaccurate and incomplete data, lack of communication between customers, item managers, and procurement personnel, and a failure to adjust management practices associated with initial procurement, on-order management, and stock retention (GAO, 2008).

The GAO recommended the Navy establish inventory performance metrics in order to track and identify areas for improvement and gauge the overall health of the Navy's

inventory management practices. The GAO specifically recommended the Navy look to establishing a metric for measuring forecast and demand accuracy on their secondary inventory as this plays a vital role in determining the fiscal requirements on part inventory levels (GAO, 2008). Establishing such metrics would enable the Navy to track and improve demand forecasting processes and better accommodate fluctuations in demand. These improvements would enable the Navy to become more efficient and also help quickly identify any deficiencies in its inventory management practices. The DoD concurred with the findings by the GAO and has begun exploring potential inventory concepts aimed to strengthen overall inventory management.

C. THESIS OBJECTIVES

This thesis seeks to establish a method to determine overall Navy inventory forecasting accuracy when compared to actual inventory demands. The creation of a forecasting accuracy method will enable the Navy to find a reference point for establishing inventory efficiency metrics and serve as the baseline for future improvements to inventory forecasting methods, thereby improving inventory management practices. This new forecasting accuracy method will allow the Navy to continually gauge the overall health of its inventory management practices. Critical outcomes are compliance with DoD regulations, improved effectiveness and responsiveness of the inventory system, and saving taxpayer dollars by reducing or eliminating excess inventory. An additional benefit of an improved inventory system would be to allow the Navy to reduce expenditures on spare part stockpiles and make those funds available to other immediate

needs. Also, improved forecasting accuracy would reduce warehouse requirements and inventory storage costs.

D. PAST THESIS WORK

1. Repair Turn-Around Time

Previous thesis studies conducted on inventory forecasting accuracy include a statistical analysis of the accuracy of the repair turn-around time (RTAT) forecast model at the Naval Inventory Control Point (Ropiak, 2000). This work sought to validate the Navy’s RTAT forecast model by comparing them to standard time series forecasting methods such as four-quarter moving average and exponential smoothing.

The study determined that assumptions implicit in the Uniform Inventory Control Program (UICP) RTAT forecast model had a significant impact on forecast accuracy. It also indicated that the Navy’s model was no more accurate than the alternative standard techniques.

This research, although closely related, is not a comparative study or validation of an established model. Rather, this thesis seeks to identify a baseline method for forecast accuracy metrics of the Navy’s entire secondary inventory.

2. Shorebased Consolidated Allowance List

A second thesis in this area of study focused on the development of an alternative demand forecasting model for Naval Aviation intermediate level inventories associated with the Shorebased Consolidated Allowance List, Yokosuka,

Japan (Onders, 1994). The model consisted of two parts. The first consists of a causal model for forecasting demand originating from aircraft carriers utilizing flying hours and carrier deployment as independent variables. The second used a time series and marginal value method to forecast causal residuals and non-carrier demand. These two parts were then combined to produce a final forecast for an individual item.

Onders (1994) differs from this study in that a comparative effort between models is not conducted in this thesis. Additionally, this study looks at both maritime and aviation secondary inventories and does not incorporate the use of causal modeling techniques.

3. Repair Turn-Around Time Modeling

Lastly, past thesis work was conducted examining the forecast accuracy for a subset of Navy inventory repair parts data utilizing a computed RTAT by the UICP (Santos, 2003). This research consisted of running bootstrap simulations with various repair part arrival rates to compare forecasting accuracy. Santos (2003) differs from this study as repair part inventory items are not the only items analyzed in this thesis for determining forecast accuracy. Additionally, this study looks at the entire population of secondary inventory vice an analysis on a subset of inventory items as was done in Santos (2003).

THIS PAGE INTENTIONALLY LEFT BLANK

II. DATA

A. SOURCES OF DATA

This thesis utilizes historical data from two Uniform Inventory Control Program (UICP) systems managed at the Naval Inventory Control Point Philadelphia (NAVICP-P) for aviation inventories and the Naval Inventory Control Point Mechanicsburg (NAVICP-M) for maritime inventories. The UICP system was developed in the 1960s to assist in materiel controls at the Naval Inventory Control Points (ICPs). The data includes both consumable and repairable inventory items.

Consumable parts consist of those items that are discarded after utilization or at the end of their usable service life. Repairable parts consist of both aviation and maritime parts that undergo a reutilization process where items are refurbished and put back into service when advantageous for financial or materiel availability purposes. Normally, repairable items are returned to serviceable condition after forwarding to designated overhaul facilities at significantly less cost and in much shorter lead times rather than going through a full procurement process.

The UICP program provides automated data processing in support the following NAVICP functions (NAVSUP Pub 542, 2001):

a. Provisioning

b. Database Maintenance

c. Stock lists and cataloging

d. Configuration management

e. Allowance

f. Load lists

g. Requisition processing

h. Transaction reporting

i. Inventory review

j. Repair and program management

k. Stratification/utilization

l. Purchase

m. Financial control

n. Accounting and disbursing

o. Simulation and research

p. Materiel readiness

q. Data Retrieval

The UICP supports the NAVICP functions of inventory review, stratification/utilization, and data retrieval for this research effort.

B. UICP

The UICP system has since undergone numerous software updates and hardware upgrades since its inception and is currently in the process of being replaced by the Navy Enterprise Resource Planning system (Navy ERP). The Navy ERP Single Supply Solution, which consolidates the wholesale

and retail supply functions of the Navy, was deployed in March of 2010 to Naval Supply Systems Command (NAVSUP).[1]

Although the UICP is a legacy system that is being phased out of NAVICP daily inventory management functions, the data remains viable for this analysis since the data contained in the UCIP systems are being utilized to populate the Navy ERP databases.

C. DATA FIELDS

For the purposes of this thesis, data was extracted from both the NAVICP-P and NAVICP-M UICP databases. Each set of historical data includes quarterly demand and forecast histories by individual National Item Identification Numbers (NIIN) and includes the following additional data fields:

FIELD	DESCRIPTION
NIIN	National Item Identification Number - a unique 9-digit numeric code identifying an item in the NATO Codification System.
COG	Cognizance Code – a two digit alphanumeric code used to identify the responsible inventory manager, the stores account and the type of materiel. A list of COG codes are found in NAVSUP P-485, Vol II, Appendix 18.
PLT	Procurement Lead-Time – overall time an item takes to be procured.
PRLT	Production Lead-Time – period between the placement of an order and receipt of the item into the supply system in quarters.
RTAT	Repair Turn-Around Time – time (in quarters)

[1] Navy ERP, http://www.erp.navy.mil/deployment_info.html

FIELD	DESCRIPTION
	an item spends in the repair system.
PRTAT	Process Repair Turn-Around Time – the amount of time an item takes to be repaired in quarters.
Replacement Price	Full cost to procure an item (in USD).
Repair Price	Price for a repairable item to be repaired (in USD).
Demand	Total number of demands for an item in a quarter.
Forecast	Forecasted number of total demands for an item in a quarter.

Table 1. Description of UICP data fields.

D DATA USED IN THIS RESEARCH

The aviation data from NAVICP-P is comprised of the fields in Table 1 for forty-one consecutive quarters of demand history by individual NIINs, covering the period December 2000 to February 2010. Additionally, the aviation data also includes twenty-one consecutive quarters of forecasting history by individual NIINs (March 2005 to March 2010). The maritime data from NAVICP-M has the same fields in Table 1 for forty consecutive quarters of demand history from June 1999 to May 2009. The maritime data also includes twenty-two consecutive quarters of forecasting history by individual NIIN from March 2005 to June 2010.

E. DATA SUMMARY

The following is a summary of the contents of both the NAVICP-M and NAVICP-P data from the UICP systems:

Type	# NIINs	# COGs
Aviation	41,005	9
Maritime	208,824	35

Table 2. Data description.

Aviation data is broken down by Cognizance Code (COG) in Figure 1. Maritime data is presented in Figure 2.

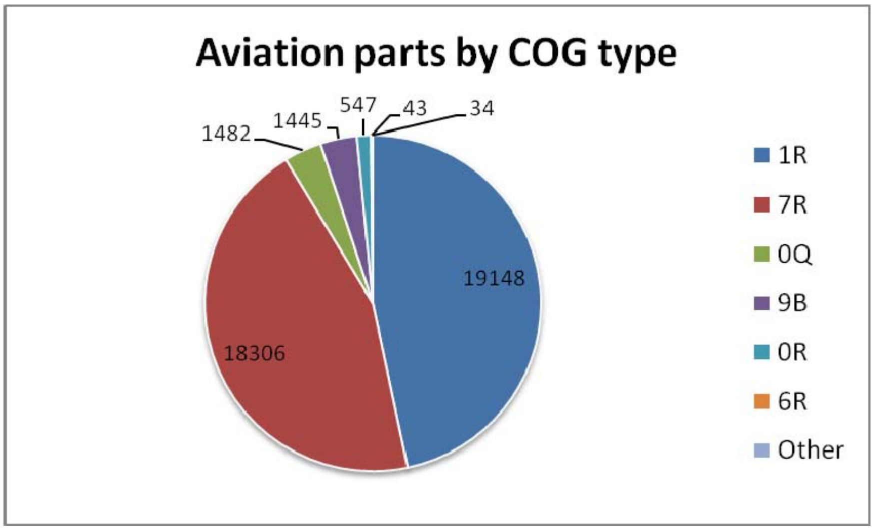


Figure 1. Distribution of aviation NIINs by COG code.

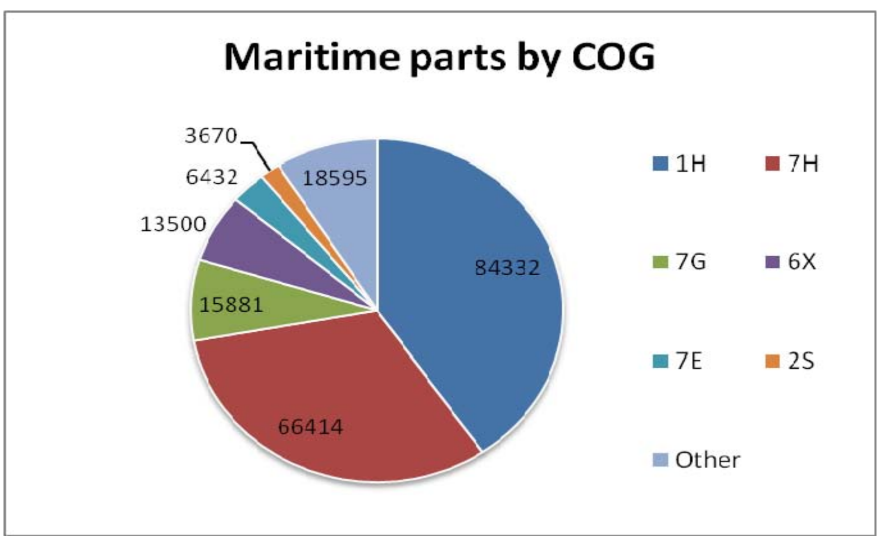


Figure 2. Distribution of maritime NIINs by COG code.

The COG code is a two-position code to identify and designate the inventory control point, which exercises supply management control. The first position of the cognizance symbol is numeric and identifies the stores account. The second position is a single-letter code that designates the inventory manager or ICP that has control of the materiel. The most commonly used cognizance symbols for the Navy are:

1R (Aviation) and 1H (Maritime) - Consumable parts

7R (Aviation) and 7H (Maritime) - Repair parts.

A complete list of COG code types and their respective inventory managers can be found in Appendix 18 of the NAVSUP P-485, Naval Supply Procedures, Volume II.

Maritime parts have a significantly larger set of COG codes compared to aviation parts. This is primarily due to the broad number of ship classes as well as the number of unique characteristics found within each ship class (e.g., DDG Flt I, DDG Flt II, DDG Flt IIA, etc.) as compared to aircraft. Maritime data taken from the NAVICP-M UICP system consisted of 208,824 NIINs from 35 different COG codes.

Table 3 presents summary statistics from the relevant fields of both aviation and maritime UICP systems.

Type	Avg Qtrly PLT	Avg Qtrly PRLT	Avg Qtrly RTAT	Avg Qtrly PRTAT	Avg Repl Price	Avg Repair Price	Avg Qtrly Demand	Avg Qtrly Forecast
Aviation	3.49	4.64	0.36	0.42	$10,426	$5,163	3.34	4.15
Maritime	3.19	4.46	0.82	0.86	$16,284	$10,627	1.50	1.55

Table 3. Summary statistics for aviation and maritime secondary inventory.

III. INVENTORY FORECAST METRICS

This chapter develops metrics useful for determining Navy inventory forecast accuracy. The goal is to provide NAVICP with metrics useful for identifying NIINs whose forecasts differ significantly from actual demand. Studies in the field of inventory management have focused on the application of time series models for measuring forecasting accuracy. Commonly used metrics include the mean absolute percentage error (MAPE) and the mean square error (MSE) (e.g., Tersine, 1998). However, this research does not use these metrics for a couple of reasons, including that they assess the fit of a single time series model to a single series, while NAVICP is most interested in evaluating forecasts in one time period over many NIINs. In addition, MAPE is not practical at NAVICP as it causes calculation errors due to division by zero. Therefore, an alternative approach is used in this research study to measure forecast error.

A. DEVELOPING PERFORMANCE INDICATORS

The simplest performance metric is just the difference between the forecast and demands over some time period. Let $F_{i,j}$ denote the forecast for NIIN *i* for period *j* and let $D_{i,j}$ denote the demand for the same NIIN and period, which could be a quarter or year. Then the performance index (PI) for NIIN *i* in period *j* is the difference between forecast and demand, expressed mathematically as:

$$PI_{i,j} = F_{i,j} - D_{i,j}. \qquad (1)$$

The problem with this type of simple performance index is that it does not provide a context within which to judge the difference. For example, the PI is the same for one NIIN that had a forecast of 10,000 items and a demand of 10,010 and another NIIN that has a forecast of 1 item and a demand of 11. That is, in both cases PI = -10, meaning the forecast was low by 10 items. However, in the first case the forecasting algorithm performed well, satisfying 10,000 out of 10,010 demands, or 99.9%, while in the second case the forecasting algorithm performed poorly, only fulfilling 1 out of 11 demands, or 9%.

This suggests Equation (1) should be modified to compare the difference between forecast and demand for time period *j* to demands for the same period, relative to the demand:

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j}}. \tag{2}$$

The performance index can now be interpreted as the fraction of demand either over-forecast (if PI is positive) or under-forecast (if PI is negative). However, while the Equation (2) performance index works in the previous example, it has two difficulties. First, when demands are 0 it is unclear what the PI value should be. Second, the calculated outputs are not symmetric in changes to the forecast versus changes to the demand. That is, for a fixed level of demand

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j}} \rightarrow -1 \text{ as } F_{i,j} \rightarrow 0$$

while

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j}} \to \infty \text{ as } F_{i,j} \to \infty.$$

On the other hand, for a fixed forecast

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j}} \to -1 \text{ as } D_{i,j} \to \infty$$

while

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j}} \to ? \text{ as } D_{i,j} \to 0.$$

The problem with the performance index being undefined when demands are zero is easily fixed by adding one to the denominator,

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1}, \tag{3}$$

but the asymmetry in performance remains and, in fact, is now more evident. That is, for a fixed level of demand

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \to -\frac{D_{i,j}}{D_{i,j} + 1} \text{ as } F_{i,j} \to 0$$

while

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \to \infty \text{ as } F_{i,j} \to \infty.$$

On the other hand, for a fixed forecast

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \to -1 \text{ as } D_{i,j} \to \infty$$

while

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \to F_{i,j} \text{ as } D_{i,j} \to 0.$$

In words, the above expressions say that with this formulation of the performance index, if the forecast is "far off" in the sense that demand is much larger than the forecast, the worst the index can be is -1. On the other hand, in the reverse case where forecast greatly exceeds demand, then the index can get arbitrarily large. This has the effect of making one type of error (forecast larger than demand) look much worse than the other type of error (demand larger than forecast).

Now, one might propose to change the denominator from demands to the forecast, as in:

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{F_{i,j} + 1}, \tag{4}$$

but this does not eliminate the asymmetry, it simply flips it into the other direction so that the error when demand is larger than forecast looks much worse than the error when forecast is larger than demand.

A solution to this dilemma is to use both Equations (3) and (4) depending on whether the forecast is greater than demand or vice versa. Specifically, when $F_{i,j} \geq D_{i,j}$ calculate the performance index as:

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \tag{5}$$

and when $F_{i,j} < D_{i,j}$ then calculate the performance index as:

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{F_{i,j} + 1}. \tag{6}$$

Equations (5) and (6) can be combined into one equation as

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \times I\left\{F_{i,j} \geq D_{i,j}\right\} + \frac{F_{i,j} - D_{i,j}}{F_{i,j} + 1} \times I\left\{F_{i,j} < D_{i,j}\right\}, \qquad (7)$$

where $I\{\bullet\}$ is the indicator function; it equals 1 when the condition in the brackets is true and it equals 0 otherwise.

The performance index in Equation (7) can be interpreted as the fraction of excess inventory compared to demand (when PI is positive) or as the fraction of unmet demand compared to the forecast (when PI is negative). Note how it behaves. For a fixed level of demand, $D_{i,j}>0$,

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \times I\left\{F_{i,j} \geq D_{i,j}\right\} + \frac{F_{i,j} - D_{i,j}}{F_{i,j} + 1} \times I\left\{F_{i,j} < D_{i,j}\right\} \to \infty \text{ as } F_{i,j} \to \infty$$

while for a fixed forecast, $F_{i,j}>0$,

$$PI_{i,j} = \frac{F_{i,j} - D_{i,j}}{D_{i,j} + 1} \times I\left\{F_{i,j} \geq D_{i,j}\right\} + \frac{F_{i,j} - D_{i,j}}{F_{i,j} + 1} \times I\left\{F_{i,j} < D_{i,j}\right\} \to -\infty \text{ as } D_{i,j} \to \infty.$$

In words, the above expressions say that, with this formulation of the performance index, if either the forecast or demand is far off then the index can get arbitrarily large in either the positive or negative direction.

To show the performance index in Equation (7) provides a symmetric measure of performance, whether the forecast is greater than demand or vice versa, is perhaps best illustrated through example. Consider the following forecasts and demands in Table 4.

	Qtr 1	**Qtr 2**
Forecast	20	9
Demand	9	20

Table 4. Hypothetical forecast and demand data by quarter.

Table 5 shows the performance indices using the asymmetric methods of Equations (5) and (6) versus the symmetric method of Equation (7). This table shows that whether the forecast is greater than demand by 9 units, as shown in Qtr 1, or if forecast is less than demands by 9 units, as in Qtr 2, Equation (7) produces results in an index of the same magnitude where sign indicates whether demands were over-forecasted (+) or under-forecasted (–).

Method	Qtr 1 performance	Qtr 2 performance
Equation (5) PI	1.10	-0.52
Equation (6) PI	0.52	-1.10
Equation(7) PI	1.10	-1.10

Table 5. Performance indices calculated on Table 4 using hypothetical data.

A second, more extreme example illustrates the asymmetry effect even more. That is, using the more divergent demand and forecast data in Table 6, Table 7 shows the performance indices using Equations (5) through (7).

	Qtr 1	Qtr 2
Forecast	120	5
Demand	5	120

Table 6. Hypothetical forecast and demand data by quarter.

Method	Qtr 1 performance	Qtr 2 performance
Equation (5) PI	19.16	-0.95
Equation (6) PI	0.95	-19.16
Equation (7) PI	19.16	-19.16

Table 7. Performance indices for Table 6 data.

Notably, the performance indices from Equations (5) and (6) again demonstrate their asymmetry whereas Equation (7) shows continued symmetry.

The asymmetry arising in Equations (5) and (6) and the symmetry of Equation (7) are even more evident when plotting actual demand and forecast data. For example, using demand and forecast data from the March through May 2006 NAVICP-P aviation data, Figure 3 plots the performance indices from Equations (5) and (6) versus demand for the same period.

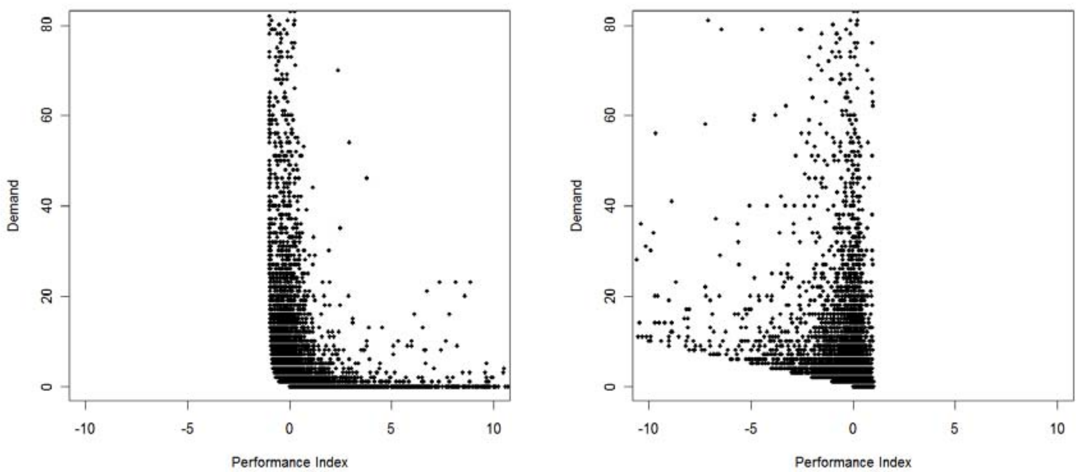


Figure 3. Performance indices versus demand for Mar-May 2006 NAVICP Aviation data for Equation (5) on the left and Equation (6) on the right.[2]

Note how in the left plot the performance indices are bounded below by -1 and in the right plot they are bounded above by +1.

[2] Note: On the right side plot, white space in the lower left region occurs due to the fact that PI in Equation (5) cannot be less than $-D_i$, which occurs when F_i=0. (the lowest possible value of F_i).

In comparison, when applying Equation (7) to the same data, Figure 4 shows the symmetry of this performance metric.

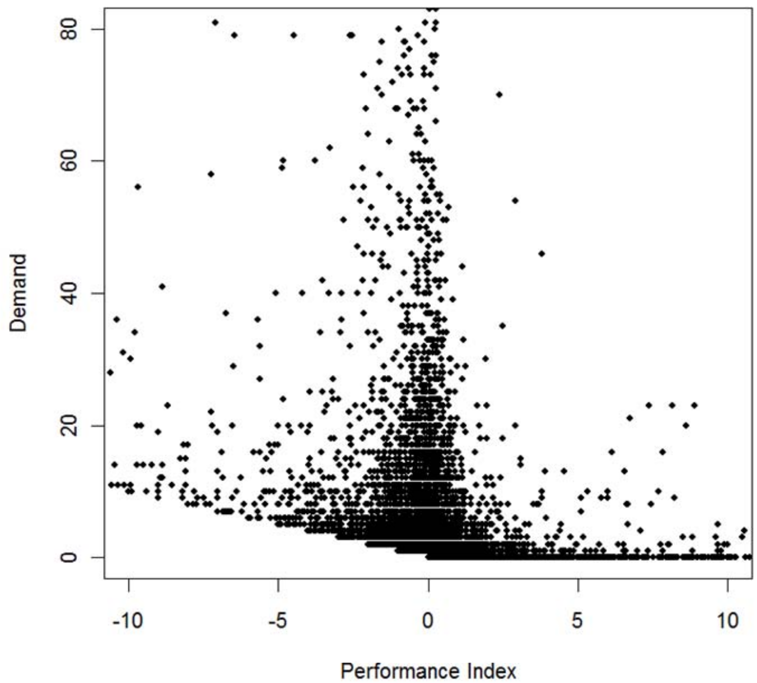


Figure 4. Performance indices for May-Mar 2006 NAVICP Aviation data using Equation (7).[3]

In a previous study conducted by NAVICP (Bencomo, 2010), a formula for measuring forecasting error (referred to as the "Lead-time Adjusted Symmetric Error: LASE") similar to Equation (7) was proposed:

$$\mathrm{LASE} = \frac{\left|F_{i,j} - D_{i,j}\right|}{\left\lceil \frac{(F_{i,j} + D_{i,j})}{2} \right\rceil + 1}. \qquad (8)$$

[3] Note: As in the right side plot of Figure 3, the white space in the lower left region occurs due to the fact that PI in Equation (7) cannot be less than $-D_i$ which occurs when F_i=0.

Although Equation (8) is a symmetric method for measuring performance, it also contains a "smoothing effect" as forecasting accuracy degrades. For example, using the hypothetical forecast and demand data from Table 8, Table 9 compares the two performance indices.

	Qtr 1	Qtr 2
Forecast	20	100
Demand	5	5

Table 8. Hypothetical set of quarterly forecast and demand.

Comparison	PI Qtr 1	PI Qtr 2
Equation (8) PI	1.11	1.78
Equation (7) PI	2.50	15.83

Table 9. Comparative PI’s on Table 8 sample data using Equation (7) and Equation (8).

By averaging forecast and demand in the denominator of Equation (8), NAVICP’s metric produces a larger smoothing effect compared to the Equation (7) performance index. Essentially, Equation (8) underestimates the magnitude of how far off the forecast is from the demand as it includes both the forecast and demand terms in the denominator of the metric.

This effect is illustrated in Table 9, where the NAVICP Equation (8) performance index increases from 1.11 to 1.78 while the Equation (7) performance index increases from 2.50 to 15.83. Notice, the forecast and demand were more divergent from quarter 1 to quarter 2 as the performance metric increase was larger in the Equation (7) calculation compared to NAVICP’s LASE calculation in Equation (8). The

smoothing effect when using NAVICP's LASE method makes identification of any large divergence between forecast and demand difficult compared to utilization of the PI calculation.

Note how Equation (8) behaves in extreme cases. For a fixed level of demand, $D_{i,j}>0$,

$$\text{LASE} = \frac{\left|F_{i,j} - D_{i,j}\right|}{\left[\dfrac{(F_{i,j} + D_{i,j})}{2}\right] + 1} \rightarrow 2 \text{ as } F_{i,j} \rightarrow \infty$$

while for a fixed forecast, $F_{i,j}>0$,

$$\text{LASE} = \frac{\left|F_{i,j} - D_{i,j}\right|}{\left[\dfrac{(F_{i,j} + D_{i,j})}{2}\right] + 1} \rightarrow -2 \text{ as } D_{i,j} \rightarrow \infty.$$

In both cases, where demand was either massively over-forecast or under-forecast, the NAVICP performance index converges to either plus two or minus two. The result is that the NAVICP metric does not have an intuitive interpretation and it makes increasingly poor forecast performance hard to distinguish.

Consider the following example in which demand is ten, $D_{i,j}=10$. If $F_{i,j}=10$ then the NAVICP performance index is 0 indicating the forecast matched the demand perfectly. However, if the forecast is double the demand ($F_{i,j}=20$), then the NAVICP performance indicator is 0.625, while if the forecast is 10 times demand ($F_{i,j}=100$) then the NAVICP performance indicator is 1.6. Finally, if the forecast is 1,000 times demand ($F_{i,j}=10{,}000$) then the NAVICP performance indicator is 1.995. Thus when the forecast got five times worse, going from $F_{i,j}=20$ to $F_{i,j}=100$ the index increased by

just under one, yet when the forecast got five *hundred* times worse, going from $F_{i,j}$=20 to $F_{i,j}$=10,000 the index increased by just over 1.3.

Also note that the NAVICP performance index does not distinguish between an under- and over-forecast. That is, the NAVICP performance index is the same whether $D_{i,j}$=10 and $F_{i,j}$=100 or $D_{i,j}$=100 and $F_{i,j}$=10. However, there are different consequences, and thus presumably different inventory management strategies and options, depending on whether an item is over- or under-stocked.

In summary, Equation (7) is the preferred performance index form as it better identifies those NIINs whose forecast and demands are significantly divergent. Further details of NAVICP's analysis can be found in Bencomo (2010).

B. QUANTIFYING AVERAGE PERFORMANCE: THE ANNUAL PERFORMANCE INDEX (API) METRIC

Measuring NIIN forecast performance on a quarterly basis is likely to be misleading as the forecasts are calculated to meet an average demand. Thus, the first metric, referred to as the Annual Performance Index (API), is based on four quarters of forecast and demand data. The API metric follows the same tenets used in the quarterly PI of Equation (7) which ensures the API is a symmetric measure regardless of whether forecast is greater than demand or if demand is greater than forecast. The main difference is that demands and forecasts are summed over four quarters:

$$API_{i,j} = \frac{\sum_{k=j}^{j-3} F_{i,k} - \sum_{k=j}^{j-3} D_{i,k}}{\sum_{k=j}^{j-3} D_{i,k} + 1} \times I\left\{\sum_{k=j}^{j-3} F_{i,k} \geq \sum_{k=j}^{j-3} D_{i,k}\right\} + \frac{\sum_{k=j}^{j-3} F_{i,k} - \sum_{k=j}^{j-3} D_{i,k}}{\sum_{k=j}^{j-3} F_{i,k} + 1} \times I\left\{\sum_{k=j}^{j-3} F_{i,k} < \sum_{k=j}^{j-3} D_{i,k}\right\}. \quad (9)$$

The choice of four quarters to sum over is based on the assumption that some NIINs might have annual cycles in their demand patterns and thus using a multiple of four quarters would minimize the effects of such a cycle. It is also motivated by the idea that, while it is desirable to smooth over some reasonable length of time in order to mitigate the effects of temporary spikes or lulls in demand, using too long of a period would have the effect of incorporating historical demand patterns and/or forecasts that may not be relevant to current inventory performance. Whether four quarters is the best choice is an open question and not studied in this thesis.

Equation (9) is based on the idea of comparing the total demands over the past four quarters to the total forecast over the same period. This seems like a sensible approach from the perspective that in each quarter there are a specific number of demands and each of those quarterly demands correspond to a particular quarterly forecast. However, an alternate approach would be to compare the forecast from a specific quarter to all of the demands that occurred in some future period of time. Implementing this requires a small modification to Equation (9). For example, if in quarter j one wanted to assess the performance of the forecast from a year ago in quarter j-3, one would just multiply that forecast times four and compare it to the demands over that same period:

$$API_{i,j} = \frac{4\times F_{i,j-3} - \sum_{k=j}^{j-3} D_{i,k}}{\sum_{k=j}^{j-3} D_{i,k} + 1} \times I\left\{4\times F_{i,j-3} \geq \sum_{k=j}^{j-3} D_{i,k}\right\} + \frac{4\times F_{i,j-3} - \sum_{k=j}^{j-3} D_{i,k}}{4\times F_{i,j-3} + 1} \times I\left\{4\times F_{i,j-3} < \sum_{k=j}^{j-3} D_{i,k}\right\}. \qquad (10)$$

This thesis will use the API from Equation (9), though NAVICP could use the multiplicative forecast approach form of Equation (10) if preferred.

C. QUANTIFYING PERFORMANCE VARIATION: THE MEDIAN ABSOLUTE DEVIATION PERFORMANCE INDEX (MPI) METRIC

Once an individual NIIN’s quarterly PI is computed using Equation (7), it may be of interest to identify NIINs whose forecasts regularly diverge from demand. The median absolute deviation (MAD) is a measure of statistical dispersion. It is preferred to other measures of variation, such as the standard deviation, as it is more resilient to outliers. For *k* observations, the MAD is calculated as

$$MAD = median\left(\left|X_1 - \tilde{X}\right|, \left|X_2 - \tilde{X}\right|, \ldots, \left|X_k - \tilde{X}\right|\right), \qquad (11)$$

where $\tilde{X} = median(X_1, \ldots, X_k)$.

For example, to compute the MAD using Equation (11), consider the following data set:

- {0, 1, 2, 3, 8}
  - The median of the data set is 2.
  - Taking the absolute deviations about 2 produces the data set {2, 1, 0, 1, 6}.

- Sorting the newly computed absolute deviations from low to high results in the following data set:

$$\{0, 1, 1, 2, 6\}$$

From this, the median is 1 and thus the MAD of the original data set is 1.

The MAD can be used to quantify the variation in quarterly performance indices in order to identify those NIINs that are regularly and routinely off in their forecasts versus those NIINs that are only intermittently off. Presumably, the forecasts in the former case are likely to be easier to correct than those in the latter case. Perhaps more importantly, the MAD will help identify those NIINs with forecasts that are on average correct, say on an annual basis, but that are routinely off on a quarterly basis, say due to churn. For purposes of this thesis, churn is identified as two quarters of over and under forecasting in a given four-quarter period. A high MAD value indicates a wide dispersion between a NIIN’s quarterly forecasts and demands. This helps in identifying NIINs with more volatility in their forecast accuracy and a NIIN’s potential for the occurrence of churn.

The use of MAD over other statistical methods, such as the standard deviation, is preferred as it is less influenced by occasional spikes or one-time anomalies in forecasting performance. For example, consider the following hypothetical quarterly performance indices for some NIIN:

$$\{1, 300, 1, 2\}$$

The standard deviation of the above performance indices is 149.3 while the MAD is 0.5. The one large performance index of 300 results in a much larger standard deviation compared to MAD. If 300 were due to a one-time demand spike, one would not want to identify the NIIN as having excess quarterly forecast variation. Rather, the measure should be insensitive to a single deviation but sensitive to two or more deviations, which is what the MAD does for a year's worth of quarterly performance indices.

To illustrate this, consider the following dataset:

{1, 300, 1, 200}

For this data, the standard deviation remains almost the same at 149.4 while the MAD significantly increases to 99.5. The standard deviation minimally changes and does not significantly differentiate between the two datasets while the MAD does.

Applying a MAD calculation to a set of quarterly PIs provides a NIIN’s forecast variability, which will be referred to as the MAD performance index or MPI. Given that the average procurement lead time for a NIIN is approximately 4.5 quarters (Bencomo, 2010), and because some NIINs likely experience an annual cycle in demand, Equation (12) shows how MPI will be calculated on a year's worth of quarterly performance index data.[4] Thus, MPI is defined as

$$MPI_{i,j} = MAD\left(PI_{i,j}, PI_{i,j-1}, PI_{i,j-2}, PI_{i,j-3}\right). \qquad (12)$$

[4] Note that as the number of quarters used in the MAD calculation is increased, the MAD requires a greater number of quarters with large performance indices before it indicates excess quarterly variation. For example, if the MAD is based on six quarters of performance index data, two or more out of the six will have to be large before the MAD is affected.

Additionally, applying a similar MAD calculation to a set of quarterly demands by NIIN provides a method that allows a differentiation between NIINs with consistent demand levels over those with intermittent demands from quarter to quarter.

D. SUMMARY

Two metrics have been developed in this chapter: API and MPI. The API measures the average forecast performance over some period of time. Herein the period is defined as a year (four quarters) but that can be modified as desired by NAVICP. The MPI measures the quarterly deviation of forecast from demand and it should provide insight into those NIINs that may be performing well on average, say according to the API, but have significant deviations quarter-by-quarter.

Adoption of these metrics will give NAVICP the ability to flag NIINs that exceed certain API and MAD thresholds for further examination and correction. Examination thresholds could be set based on staff workload capabilities, item cost, or COG code. Such an approach gives NAVICP the capability to focus attention on those NIINs that demonstrate poor forecasting performance, enabling improvements to the overall inventory management system.

IV. APPLICATION OF THE METRICS

This chapter illustrates the application of the API and MPI to inventory forecast evaluation. In the first section, hypothetical NIIN demand and forecast data are compared using the API and MPI metrics to provide some insight into how the metrics work and what they indicate about forecast performance. The second section then demonstrates how to apply the metrics to a complete inventory to identify actual NIINs that have problematic forecasts.

A. ILLUSTRATING HOW THE METRICS WORK

To illustrate how the metrics work, consider the four hypothetical NIINs in Table 10, and their demand and forecast history from June 2005 to May 2006. These are NIINs with forecasts that are "well behaved" in the sense that the forecasts closely match the demands quarter by quarter.

Jun - May Forecast and Demand quantities by quarter								
	Jun-Aug		Sept-Nov		Dec-Feb		Mar-May	
NIIN	Forecast	Demand	Forecast	Demand	Forecast	Demand	Forecast	Demand
012345678	5	6	6	6	5	6	6	7
123456789	0	0	0	1	0	0	0	0
234567890	10	8	8	9	11	12	12	10
345678901	0	0	0	0	0	0	0	0

Table 10. Quarterly forecast and demand for four hypothetical NIINs.

Using Equation (7), Table 11 gives the API metrics for the four NIINs.

NIIN	API
012345678	-0.13
123456789	-1.00
234567890	0.05
345678901	0.00

Table 11. API metrics for NIINs in Table 10.

Here we see relatively small API indices, which we interpret as forecasts and demands closely matching from quarter to quarter. The largest magnitude difference is NIIN 123456789, where the -1.0 indicates that the NIIN was 100% under-forecast over the four quarters (the forecast was zero for all four quarters while there was a demand of one). In contrast, NIIN 012345678 actually under-forecast by three units, but that was out of 25 demands, so on a percentage basis the forecast for this NIIN actually performed better than the previous NIIN. Of course, one might argue that because of the low demand NIIN 123456789 is of little interest, but as will be shown later in this section, such NIINs can be identified by plotting API versus demand.

Using Equation (12), Table 12 gives the MPI metrics for the four NIINs.

NIIN	MPI
012345678	0
123456789	0
234567890	1
345678901	0

Table 12. MPI metrics for NIINs in Table 10.

As MPI quantifies a NIIN's forecast variability from quarter to quarter. The small MPI values in Table 12

indicate that demand and forecast performance numbers remain relatively consistent quarter to quarter. This is validated in Table 10 where the NIIN forecast and demand numbers remain relatively close to each other in at least two or more quarters for each NIIN.

Now, consider another example in Table 13, where the hypothetical NIINs have very poor forecasts showing churn.

Jun - May Forecast and Demand quantities by quarter								
	Jun-Aug		Sept-Nov		Dec-Feb		Mar-May	
NIIN	Forecast	Demand	Forecast	Demand	Forecast	Demand	Forecast	Demand
912345678	5	504	1500	4	6	5000	4000	1
923456789	3	200	15	3500	3300	22	1500	900
934567890	50	50	100	1	500	2	1	580
945678901	1000	1	1	80	800	300	1	1500

Table 13. Quarterly forecast and demand for four hypothetical NIINs with churn.

It becomes evident upon examination of the computed APIs and MPIs in Table 14 and Table 15 that churn exists because API is low but MPI is very high.

NIIN	API
912345678	0.00
923456789	0.04
934567890	0.03
945678901	-0.04

Table 14. API Metrics for the NIINs in Table 13.

NIIN	MPI
912345678	506.31
923456789	95.89
934567890	83.00
945678901	269.50

Table 15. MPI Metrics for the NIINs in Table 13.

Most notably, the significantly large MPI numbers are due to the excessively large variation and significant swings in over- and under-forecasting performance from quarter to quarter for each NIIN in Table 13. However, when the demand and forecast numbers are aggregated to calculate their respective API's, the NIINs have very small API indices because their four quarter forecast and demand sums are very close.

In yet another example, Table 16 shows a set of hypothetical NIINs whose quarterly forecast and demand data reflect two patterns. The first two NIINs in Table 16 show intermittent demand or forecast spikes while the last two NIINs show consistently high under- and over-forecasting for at least two or more quarters. The resulting API and MPI calculations are provided in Tables 17 and 18.

Jun - May Forecast and Demand quantities by quarter								
	Jun-Aug		Sept-Nov		Dec-Feb		Mar-May	
NIIN	Forecast	Demand	Forecast	Demand	Forecast	Demand	Forecast	Demand
812345678	50	50	25	24	8000	2	4	4
823456789	3	5000	8	10	44	45	15	16
834567890	0	5000	0	1000	1	1	1	0
845678901	5000	0	1000	0	600	0	1000	0

Table 16. Quarterly forecast and demand for hypothetical NIINs with spikes in demand or forecast and large variability.

NIIN	API
812345678	98.75
823456789	-70.44
834567890	-1999.70
845678901	7600.00

Table 17. API metrics for NIINs in Table 16.

NIIN	MPI
812345678	0.02
823456789	0.10
834567890	500.50
845678901	200.00

Table 18. MPI metrics for NIINs in Table 16.

As shown in Tables 17 and 18, intermittent spikes in quarterly demand or forecasting quantities produce a large API but low MPI indices. The low MPI is due to the MAD, which is less sensitive to the occasional outlier such as a one-time spike in demand. However, if forecasting or demand quantities are significantly off for two or more quarters in a four-quarter period, as demonstrated by the last two NIINs in Tables 17 and 18, they produce both high API and MPI indices.

B. APPLYING THE METRICS TO ACTUAL DATA

Plotting the MPI and API indices for NIINs allows for a graphical depiction of NIIN performance. Figure 5 demonstrates such a graph for NAVICP-P data from the period June 2005 to May 2006 where MPI is plotted on the y-axis and API is plotted on the x-axis).

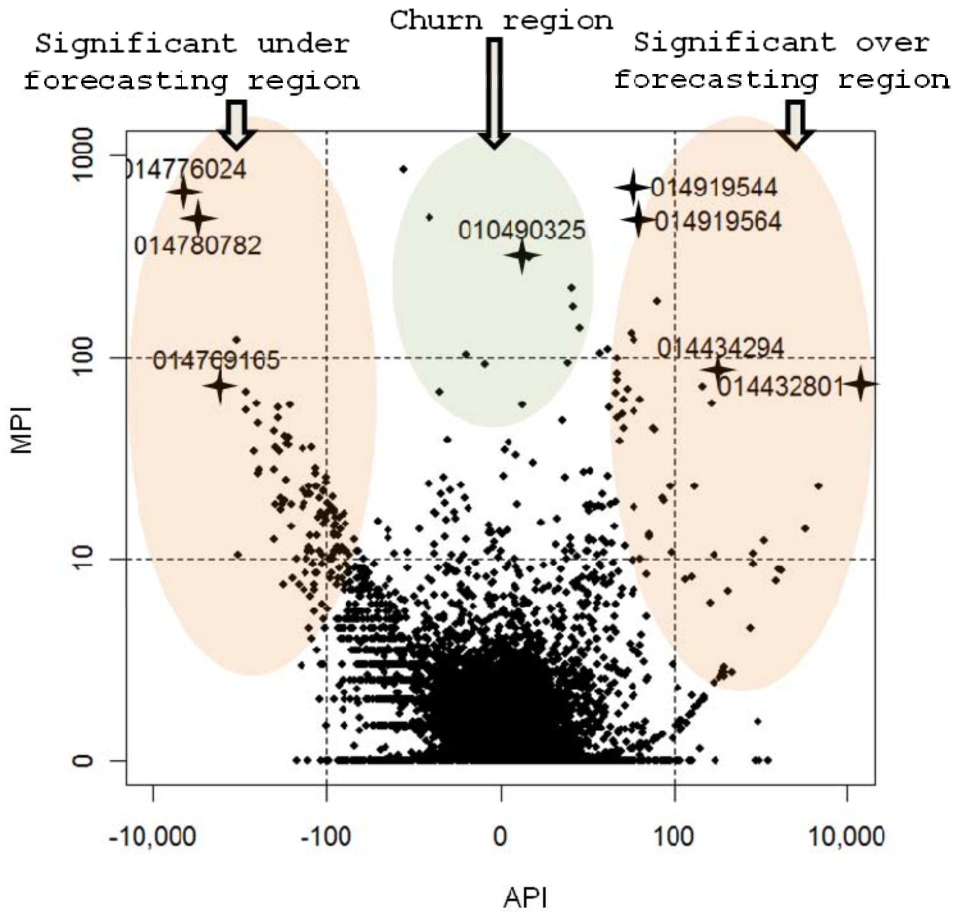


Figure 5. Plot of API vs. MPI for NAVICP-P NIINs for the period June 2005 – May 2006.

In Figure 5, the highlighted regions that depict areas where NIINs are significantly under- or over-forecast. In addition, the churn region is identified in Figure 5 where, for example, NIIN 010490325 is identified. Seven additional NIINs, three in the under-forecasting region and four in the over-forecasting region, are also identified. Table 19 then shows the actual forecast and demand data where it is clear that NIIN 010490325 experiences churn with three quarters of significant over-forecasting and one quarter of significant under-forecasting.

June 2005 – May 2006 Forecast and Demand quantities by quarter[5]									
		Jun-Aug		Sept-Nov		Dec-Feb		Mar-May	
NIIN	COG	Forecast	Demand	Forecast	Demand	Forecast	Demand	Forecast	Demand
014776024	1R	0	507	0	467	0	1850	0	1773
014780782	1R	0	194	0	1876	0	0	0	952
014769165	1R	0	303	0	449	0	423	0	710
010490325	7R	739	0	649	1056	649	0	649	396
014919544	1R	1441	28	1441	0	1441	0	1441	130
014919564	1R	957	0	11	0	957	70	957	0
014434294	1R	6	0	175	0	175	0	0	0
014432801	0Q	3345	0	3790	0	3790	0	3931	0

Table 19. Quarterly Forecast and Demand quantities for NIINs listed in Figure 5.

Additionally, notice the NIINs identified in the highlighted under-forecast region in Figure 5, and compare them against their corresponding forecast and demand quantities in Table 19. Here we see those NIINs residing in the under-forecast region are indeed experiencing significant under-forecasting as demonstrated by the consistently high demand numbers in each quarter with zero forecast quantities. Conversely, if we look at NIIN’s residing in the over-forecasting region as identified in Figure 5, you see the opposite effect, as NIINs in this case contain significantly large forecasting quantities against zero demands.

This methodology can also be selectively applied by the subsetting to particular NIINs of interest, say by COG code or perhaps by price (for example, with NIINs over a certain cost threshold). Figure 6 shows the resulting plots for NIINs separated by consumable and repairables for June 2005 through May 2006 aviation data. Notably, repairable NIINs

[5] Forecasting numbers rounded to nearest whole number.

located on the right side plot of Figure 6 exhibit a much tighter grouping over the consumable NIINs in the left plot suggesting that the repairables as a group have better forecasts than the consumables.

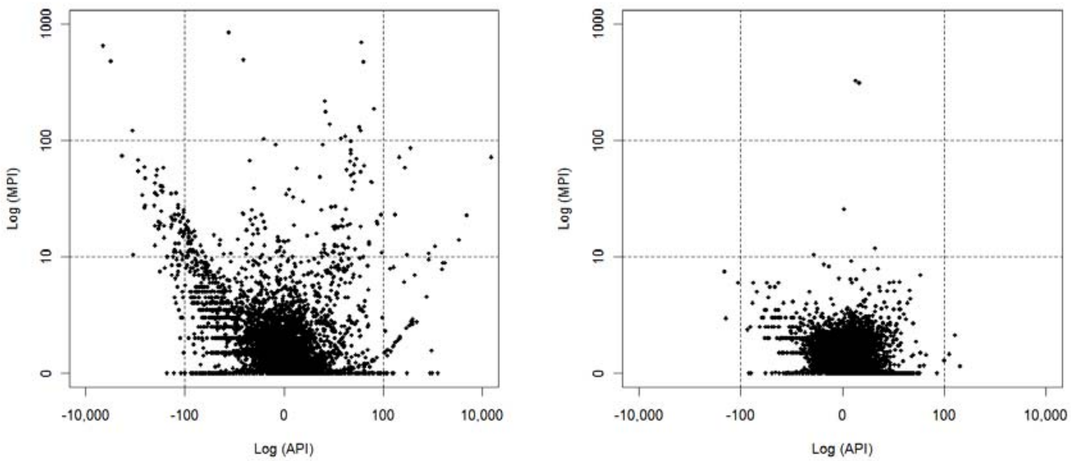


Figure 6. Comparison between consumable aviation NIINs on left against repairable aviation NIINs on right for the period June 2005 – May 2006.

Similarly, Figure 7 shows a comparison of NIINs that cost less than $10,000 versus NIINs that cost $10,000 or more.

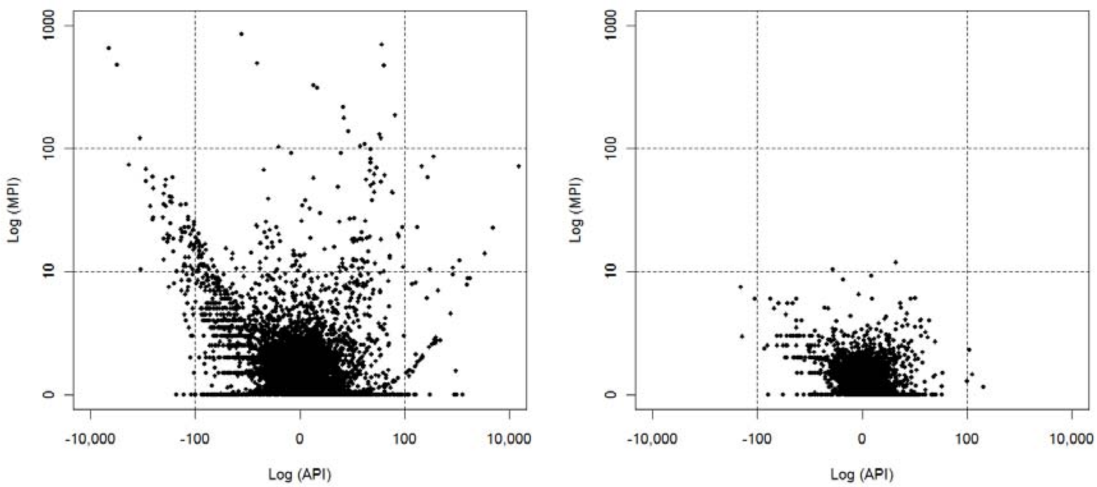


Figure 7. Aviation NIINs that cost less than $10K on left compared against aviation NIINs that cost $10K or more on right for the period June 2005 – May 2006.

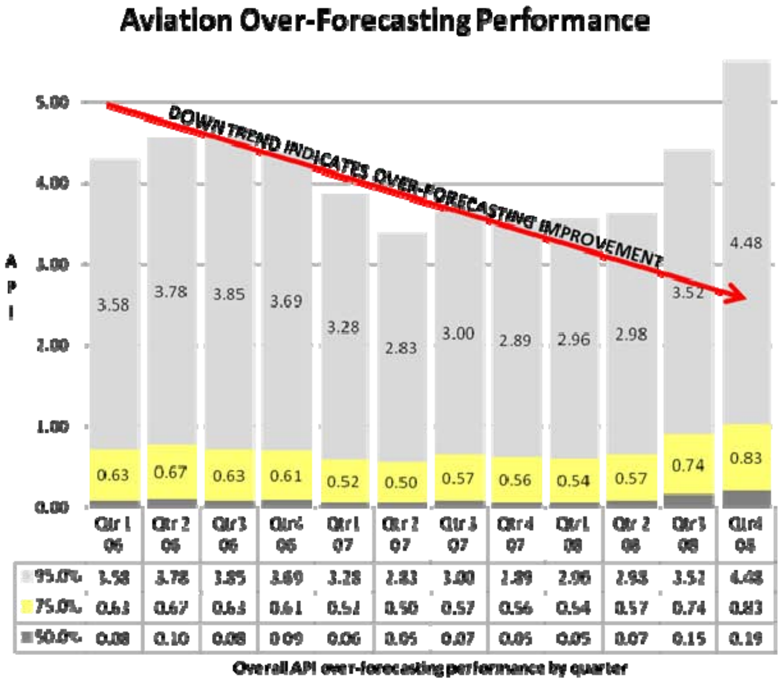
Note that for the NIINs costing $10K or more show better performance than those costing less than $10K in the sense that the left side plot in Figure 7 is larger and more dispersed when compared to the right side plot in Figure 7. This suggests that NAVICP-P appropriately places more scrutiny on forecasting accuracy on the more expensive NIINs.

That said, the right side plot is still useful for identifying expensive NIINs that are not performing as well as equivalent NIINs by simply tightening the thresholds and flagging NIINs whose MPI and API performance indices are outside the norm for expensive NIINs. Identifying those NIINs with poor forecasting performance in this fashion, so they can be audited by NAVICP personnel, should facilitate incremental forecasting performance improvements.

The API metric can also be used to summarize overall inventory performance. This would allow NAVICP to capture overall forecasting performance for all secondary inventories and determine if forecasting performance is improving or degrading over time. It is accomplished by graphically displaying the API percentiles by quarter over time, where the median, 75th percentile, and 95th percentiles for the API metrics for the over- and under-forecasted NIINs are plotted by quarter.

Figure 8 illustrates the idea using the NAVICP-P aviation data from 2006 to 2008. The top plot is for all the NIINs with over-forecasts and improving forecasts across the entire inventory would be visible as a downward trend in the bars over time. The bottom plot is for all the NIINs with under-forecasts and improving forecasts across the entire inventory would be visible as an upward trend in the bars over time.

The top of the grey bar indicates the 95th percentile of API for the particular quarter. This means that 95 percent of all the aviation NIIN APIs were less than or equal to that value. Similarly, the yellow and dark gray bars show the 75th and 50th percentiles, respectively. Therefore, in the first quarter of 2006, 95% of all over-forecast aviation NIINs had APIs less than 3.58, 75% had APIs less than 0.68, and 50% had APIs less than 0.08. Thus, 50% of the over-forecast NIINs have excellent API metrics (less than 0.08). On the other hand, 5% of the NIINs have APIs greater than 3.6, which means those NIINs have forecasts that are more than 3.5 times the number of demands.



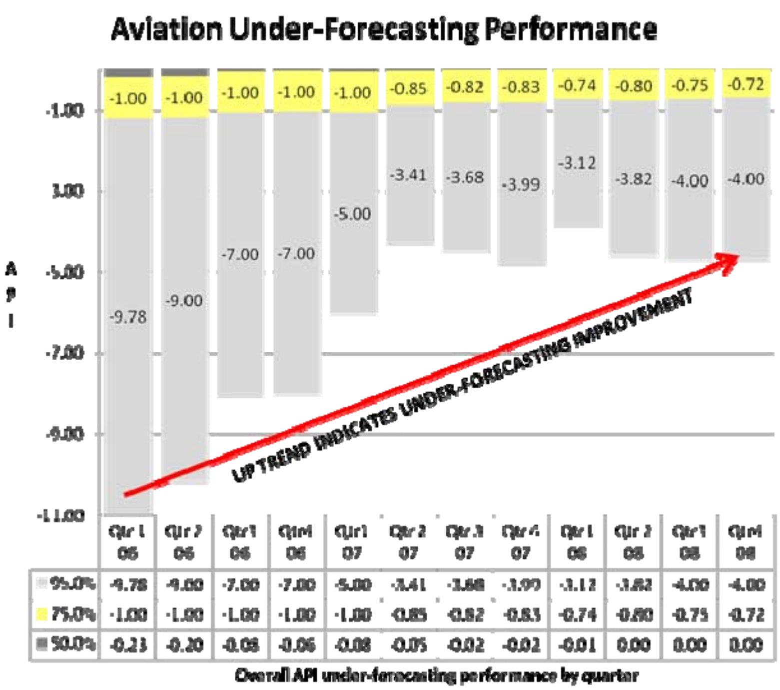


Figure 8. Quarterly API statistical percentiles on NAVICP-P aviation data from 2006 to 2008.

Using Figure 8, NAVICP can determine whether forecasting performance is improving or degrading over time as depicted by the trend line in both charts. A trend line that tends towards zero indicates improvement in forecasting performance over time. This method allows NAVICP to generate a graphical illustration for showing overall forecasting performance of secondary inventory.

C. SUMMARY

Application of the API and MPI metrics and their utility has been show in this section. Incorporating the use of the API and MPI metrics provides NAVICP with a method to measure forecasting performance. This method allows for the identification of NIINs that exceed certain established thresholds for further examination and correction. Examination thresholds could be based on, but not limited to, staff workload capabilities, item cost, or COG codes. Lastly, a method using the API indices is demonstrated and allows NAVICP the capability to capture overall quarterly forecasting trends based on the statistical spread of the quarterly API. This provides NAVICP with insight into whether forecasting performance is improving or degrading over time.

V. CONCLUSIONS

This research establishes metrics for determining overall Navy secondary inventory forecasting accuracy when compared to actual demands at the Naval Inventory Control Point (NAVICP). Specifically, two performance metrics were introduced: the API and the MPI. API measures forecasting accuracy of secondary inventory when compared against demand or forecast performance over a four-quarter period. MPI measures the quarterly variability of forecast errors over the same period.

The API and MPI metrics allow for the identification of poorly forecast NAVICP secondary inventory items. The metrics can be applied to entire inventories or subsets of items based on type, demand, or cost. In addition, the API metric can be used to show overall inventory performance, providing NAVICP with a graphical means to assess forecasting performance improvements (or degradations) over time.

The new forecasting accuracy methods developed in this research will allow the Navy to continually gauge the overall health of their inventory management practices and provide a method for improving forecasting accuracy. Additionally, they will assist NAVICP in complying with DoD directives that require NAVICP to monitor and continually develop improvements to inventory management practices (DoD, 2004).

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Bencomo, L.A., *Demand Forecast Accuracy Metric: LASE (Lead-time Adjusted Symmetric Error),* Naval Inventory Control Point, NAVICP Code 04, September 30, 2010.

Department of the US Navy Fleet Material Support Office, *UICP Executive Handbook*, NAVSUP Pub 542, September 2001.

DoD 4140.1R, *DoD Supply Chain Materiel Management Regulation*, pg. 208, May 23, 2003.

DoD Directive 4140.1, *Supply Chain Materiel Management Policy,* April 22, 2004.

NAVSUP Publication 485, *Naval Supply Procedures Volume II – Supply Appendices,* Appendix 18, May 18, 2009.

Navy ERP Program Manager, (2011). NAVY ERP Program. Retrieved on February 17, 2011, from http://www.erp.navy.mil/deployment_info.html

Onders, Randal J., *A Demand Forecasting Model for a Naval Aviation Intermediate Level Inventory – The Shorebased Consolidated Allowance List (SHORCAL), Yokosuka, Japan.* Master’s Thesis. Naval Postgraduate School. Monterey, CA, December 1994.

Ropiak, Michal J., *Statistical Analysis of the Naval Inventory Control Point Repair Turn-Around Time Forecast Model*. Master’s Thesis. Naval Postgraduate School. Monterey, CA, June 2000.

Santos, William O., *An Analysis of the Prediction Accuracy of the U.S. Navy Repair Turn-Around Time Forecast Model.* Master’s Thesis. Naval Postgraduate School. Monterey, CA, June 2003.

Tersine, Richard J., *Principles of Inventory and Materials Management,* 4th edition, Prentice Hall, Upper Saddle River, NJ, 1998.

United States Government Accountability Office (GAO), *Defense Inventory, Management Actions Needed to Improve the Cost Efficiency of the Navy's Spare Parts Inventory,* GAO-09-103. December 2008.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Walt DeGrange, CDR, SC, USN
   Director, Operations Research & Systems Support Mech.
   Naval Inventory Control Point Mechanicsburg
   Mechanicsburg, Pennsylvania

4. Ronald D. Fricker, Jr., Associate Professor
   Naval Postgraduate School
   Monterey, California

5. Patrick A. R. Burson, CDR, SC, USN
   Naval Postgraduate School
   Monterey, California

6. Kenneth J. Jackson, LCDR, SC, USN
   USS RONALD REAGAN (CVN 76)